

CONFERENCE PROGRAM

Modern Advances in Assessment: Testing and Digital Technology, Policies, and Guidelines

July 3-5, 2012

July 2, 2012 (*Pre-Conference Workshops*)

The Royal Tropical Institute, Amsterdam, The Netherlands

Conference website: www.itc2012ams.com



ISO10667: A Standard for Assessment in Work and Organizational Settings

Chair: Dave Bartram, SHL Group Ltd., (UK)

Panellists

Anders Sjöberg (Stockholm University, Sweden)

Marise Born (Erasmus University Rotterdam, Netherlands)

Wayne Camara (College Board, USA)

Sverre Nielsen (Norwegian Psychological Association, Norway)

This panel will highlight the major aspects of the standard and explore their implications for assessment work. Specifically, the panel will describe the content of ISO 10667 and discuss its major implications. The chair will provide a briefing outlining the background to the development of the standard and its structure. Each panellist will provide an overview of the contents of a section and highlight the implications for practice. The chair will guide the panel and the audience in an exploration of ways in which the standard might be used within our practice as assessment providers to educate clients and improve quality. The audience will be given the opportunity to ask questions and participate in the discussion of implications. The session will close with an exploration of the opportunities the new standard creates.

Short scales for psychological research – applicability, benefits, and potential limitations

Chairs: Christoph J. Kemper (GESIS - Leibniz Institute for the social sciences, Germany) & Beatrice Rammstedt (GESIS - Leibniz Institute for the social sciences, Germany)

Symposium abstract

In recent years, the demand of short measures for psychological constructs has been growing, e.g. in studies conducted online or large-scale surveys. If online questionnaires are too lengthy, break-off and item nonresponse may increase substantially leading to a lower quality of research data. Due to severe monetary and time constraints, measures used in large-scale surveys have to be short but still reliable and valid indicators of a psychological construct. In this symposium, the construction and validation of three short scales measuring diverse personality and ability constructs will be presented. The fourth presentation will focus on the delicate task of short scale construction. It reviews different approaches and proposes guidelines for the construction process in order to ensure that short scales are not only economic but also sufficiently reliable and valid. Different aspects of psychometric short scales will be discussed, e.g. their applicability in different contexts, their benefits, and potential limitations.

Paper 1***Are you optimistic or pessimistic, or both? - The LifeOrientationTest-Revised***

Markus Zenger (University of Leipzig, Germany)

Andreas Hinz (University of Leipzig, Germany)

Yve Stöbel-Richter (University of Leipzig, Germany)

Elmar Brähler (University of Leipzig, Germany)

The Life Orientation Test-Revised (LOT-R) was developed to measure the construct of optimism, defined as generalized outcome expectancy. Originally designed as a uni-dimensional scale with the two antipoles optimism and pessimism, several studies based on specific samples could show that optimism and pessimism can be seen as distinct dimensions. With data derived from a representative sample of the German general population ($N = 2,372$, age 18-93y.), the factorial structure of the LOT-R was tested using confirmatory factor analysis. Several model fit indices indicate that the assumption of a bi-dimensional structure of the LOT-R fits the data much better than the uni-dimensional structure. Besides, the LOT-R can be applied in clinical settings to predict psychosocial outcomes. In a study with 387 cancer patients, LOT-R scores significantly predicted anxiety, depression and health related quality of life three months later. The predictive power of optimism and pessimism differed between males and females. Furthermore, the brevity of the questionnaire (three items for optimism and pessimism, respectively, and four filler items) allows an economical assessment of optimism and pessimism for individual screening purposes as well as for epidemiological research.

Paper 2

Brief knowledge scales for the measurement of crystallized intelligence

Stefan Schipolowski (Humboldt-Universität zu Berlin, Germany)

Oliver Wilhelm (Ulm University, Germany)

Ulrich Schroeders (Humboldt-Universität zu Berlin, Germany)

Anastasiya Kovaleva (GESIS - Leibniz Institute for the social sciences, Germany)

Christoph J. Kemper (GESIS - Leibniz Institute for the social sciences, Germany)

Crystallized intelligence (*gc*) is a well-established cognitive ability factor that has been conceptualized as reflecting influences of learning, education, and acculturation. Although *gc* is best measured by capturing knowledge from different domains, in practice *gc* assessments are often limited to verbal skills (e.g., vocabulary). Based on a large item pool covering declarative knowledge from 16 domains in natural sciences, the humanities, and civics we compiled a 32 item *gc* assessment that can be completed in approximately 10 minutes and includes 2 items from each knowledge domain. Based on a sample of 1,100 German adults covering a broad age range and various educational backgrounds we derive an ultra-short version for 5 minutes testing time. First, we investigate mean, floor and ceiling effects as a function of context variables. Second, we compare competing measurement models in terms of model fit and reliability of the latent factors. Third, locally weighted measurement models will be used to evaluate the relationship of *gc* with age and educational background. These analyses are extended with an examination of mean trajectories and differentiation-differentiation of factor space. Potential and problems of contemporary ability measurement concepts for survey research will be discussed.

Paper 3

The Vocabulary and Overclaiming Test (VOC-T)

Matthias Ziegler (Humboldt-Universität zu Berlin, Germany)

Christoph J. Kemper (GESIS - Leibniz Institute for the social sciences, Germany)

Beatrice Rammstedt (GESIS - Leibniz Institute for the social sciences, Germany)

The present research aimed at constructing a questionnaire measuring overclaiming tendencies (VOC-T-bias) as an indicator of self-enhancement. An approach was used which also allows to estimate a score for vocabulary knowledge, the accuracy index (VOC-T-accuracy) using signal detection theory. For construction purposes, an online study was conducted with $N = 1,176$ participants. The resulting questionnaire, named Vocabulary and Overclaiming – Test (VOC-T) was investigated with regard to its psychometric properties in two further studies. Study 2 used data from a population representative sample ($N = 527$) and Study 3 was another online survey ($N = 933$). Results show that reliability estimates were satisfactory for the VOC-T-bias index but less so for the VOC-T-accuracy index. Overclaiming did not correlate with knowledge but was sensitive to self-enhancement supporting the test-score's construct validity. The VOC-T-accuracy index in turn covaried with general knowledge and more so with verbal knowledge also supporting construct validity. Moreover, the VOC-T-accuracy index also had a meaningful correlation with age in both validation studies. All in all, the psychometric properties can be regarded as sufficient to recommend the VOC-T for research purposes.

Paper 4

Multi-item scales: Quality criteria and reduction of length

Hilde Tobi (Wageningen University, The Netherlands)

Ynte van Dam (Wageningen University, The Netherlands)

There are several reasons to reduce scale length and there are at least as many approaches to scale length reduction being used. The literature, however, gives little guidance as to how to reach the best possible short version. The aim of this paper is to come up with some guidance based on the quality criteria: validity, reliability and discriminative power. An inventory of methods is presented and measurements to assess a particular quality criterion. Each of these methods and measurements seems to favor or discard particular item characteristics. How this results in a particular short version is illustrated with the Connected to Nature Scale (Mayer & Frantz, 2004) and the NEP-scale which aims at the construct "beliefs" within the Value-Belief-Norm theory (Dunlap, van Liere, Mertig, & Jones, 2000).

Discussant

Johannes Lutz (University of Potsdam, Germany)